



# **“Big Data” in NASA and Beyond**

*L. Harper Pryor - CSC*

NAS Division

NASA Ames Research Center

[harper.pryor@nasa.gov](mailto:harper.pryor@nasa.gov)

2012 Summer Short Course for Earth System  
Modeling and Supercomputing



# **“Big Data”**

What is it?

Why do you care?

## **NASA’s Big Data**

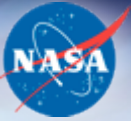
What's out there?

Where is it?

How do you access/acquire it?



# Taming the Big Data Beast



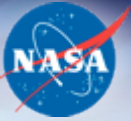
Will it be this...



Or this...



# Your Speaker – L. Harper Pryor



CSC's Program Manager for support to NAS

Over 30 years at GSFC supporting major programs in computation and science data processing for earth and space science missions

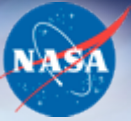
- Helped transform the NCCS into the NASA Center for Climate Simulation, with an emphasis on data intensive computing
- Coordinated support to major scientific field programs and provided system engineering to high profile projects such as the MAP '05 Hurricane Tracking Project
- Provided System Engineering and technical leadership for computing support to NASA's Global Modeling and Assimilation Office
- Supported many NASA flight and instrument programs, including the Solar Maximum Mission (SMM), Shuttle Solar Backscatter Ultraviolet experiment (SSBUV), Cosmic Background Explorer (COBE), Tropical Rainfall Monitoring Mission (TRMM), Landsat Data Continuity Mission (LDCM), and the Global Precipitation Measuring Mission (GPM)

# Acknowledgements



Grateful thanks to the following sources for information used on various of the slides that follow...

- CSC ClimateEdge presentation
- NSF Earthcube presentation
- NASA EOSDIS documentation and websites
- GES DISC training materials
- OSTP press releases and backgrounders



# What is the Big Data Problem?

The explosive growth of massive datasets, termed "Big Data", is fueled by pervasive devices observing the natural world, high-fidelity scientific instruments, and mobile and online sensors embedded in our daily lives.

This data deluge exhibits not just volume and velocity but also variability and diversity of structure, completeness and domain.

Big Data offers an unprecedented opportunity to understand scientific and social processes, and for deep insight that transcends situational awareness.

## **However...**

The potential impact of these data is constrained by our ability to rapidly and comprehensively navigate and analyze them

Their size overwhelms traditional data management and analysis techniques, and requires novel algorithms, infrastructure and frameworks to support advanced analytics

# Government Agencies Face New Challenges Associated with Climate Change “Big Data”



## Explosive Growth in Data

- Big and getting bigger: ~many tens of Petabytes today and growing exponentially
- Satellite observations
- Climate model results
- Other sensors

## New government mandates

- Open Government initiatives
- “Platform / Apps” mentality – get the data into the public domain

## Shrinking budgets

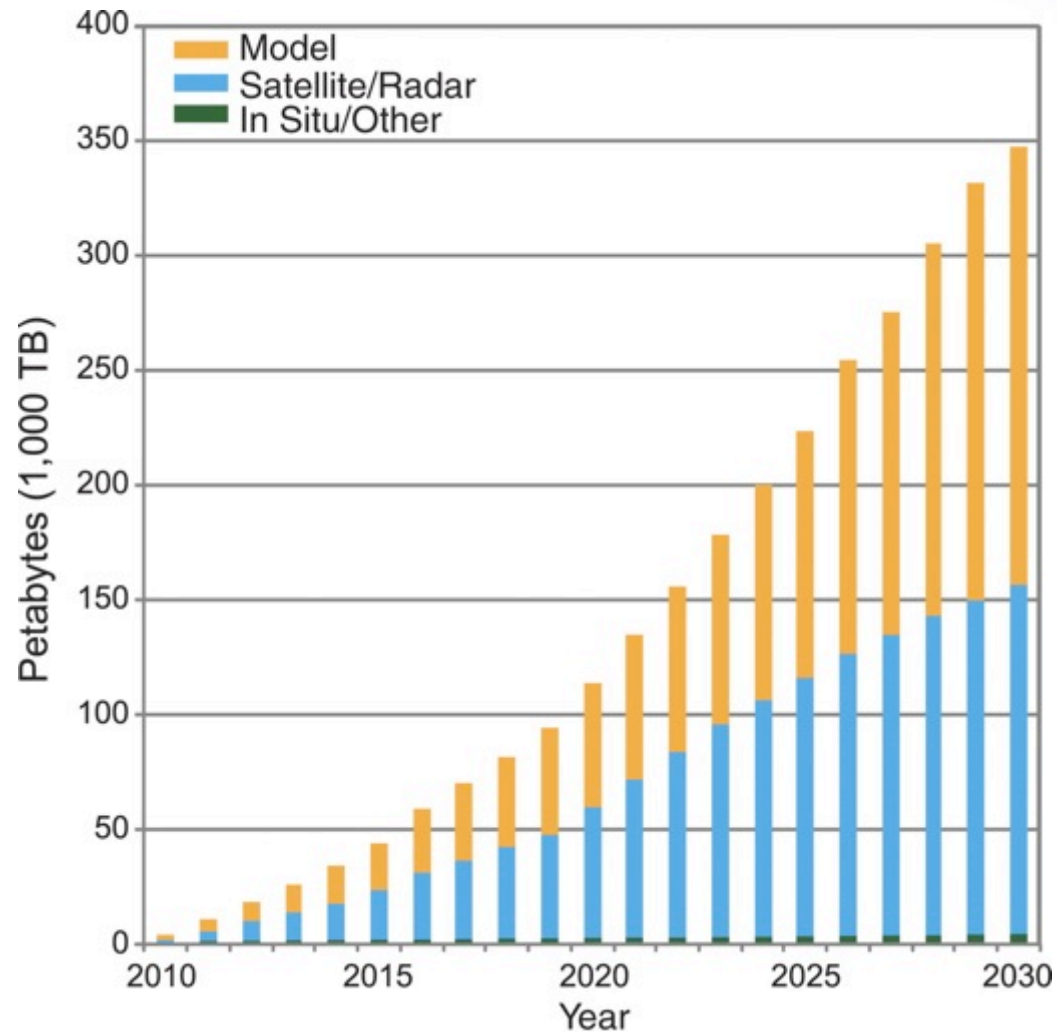
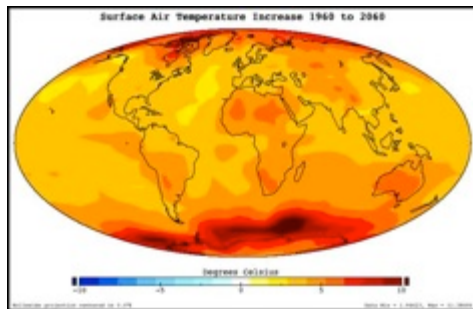
- Do more with less
- Must attract new, non-traditional consumers of the data in order to justify the existence of the data collection platforms

Ultimate goal:  
enable the  
extraction of  
*actionable*  
*information*  
from the data





# Climate Change Data Alone is Projected to Grow to Nearly 350 Petabytes by 2030





# Big Data is both Size and Speed

## Modern geoscience

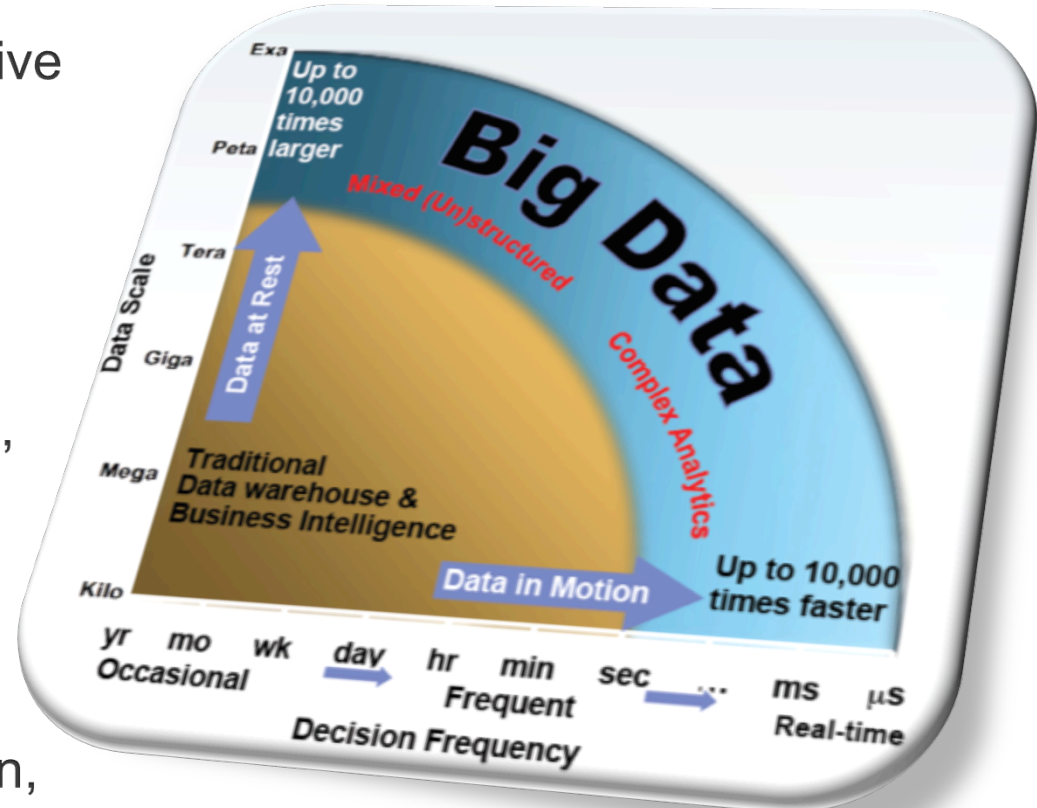
- Data- and compute-intensive
- Integrative, multi-scale

## Multi-disciplinary collaborations to address complexity

- Individuals, groups, teams, communities

## Sea of Data

- Age of Observation
- Distributed, central repositories, sensor- driven, diverse, etc.



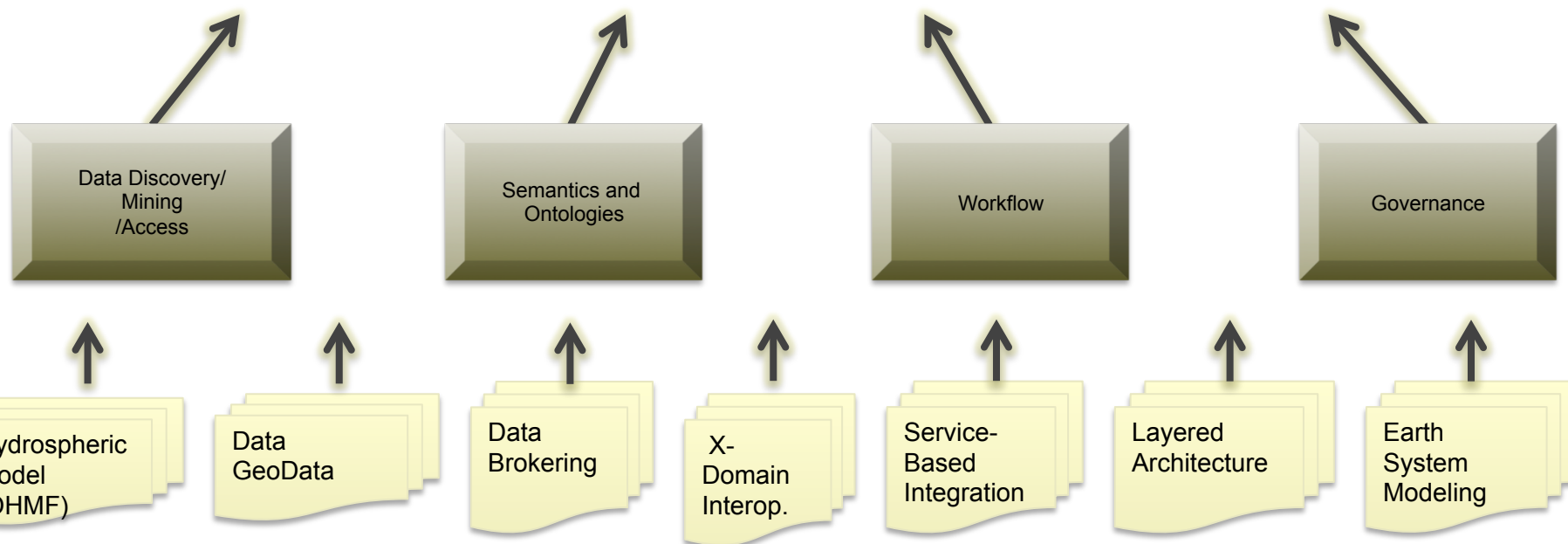
# There is more to Big Data than just data!



## EarthCube

### National Data Infrastructure for Earth System Science

NSF sponsored effort to collaboratively produce a framework to form an integrated & synergistic path forward to a data and knowledge management system to support 21<sup>st</sup> Century science



# So who is doing something about this?



Everyone!!

Academia – IT Industry – Business - Government

Myriad efforts throughout Government  
Programs in just about every department/agency  
“Big Data Research and Development Initiative”  
-- an overarching administration initiative



# Big Data Research and Development Initiative

- Announced by the Obama administration in March 2012
- Six Federal departments/agencies, \$200 million in new commitments
- To “greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data”
- Coordinated by OSTP Senior Steering Group on Big Data

***“In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education, and national security”***

***Dr. John P. Holdren, Assistant to the President and  
Director of the White House Office of Science and Technology Policy***

# Big Data Research and Development Initiative

**NSF and NIH** - Joint solicitation for Core Techniques and Technologies for Advancing Big Data Science & Engineering

**NSF** - Methods to derive knowledge from data; infrastructure to manage, curate, and serve data; Approaches to education and workforce development

**NIH** - 1000 Genomes Project Data on Cloud

**DOD** - Data to Decisions - ~\$250 million annually across military departments and ~25 million annually in the DARPA XDATA program

**DOE** - Establish the Scalable Data Management, Analysis and Visualization (SDAV) Institute within the SCIDAC program

**USGS** – Big Data for Earth System Science

**This is just a piece of the puzzle - there are myriad other efforts throughout the Government!**

# Big Data Initiatives and Programs in NASA



**Advanced Information Systems Technology (AIST)** – Reduce the cost of evolving NASA information systems to support future Earth Observation missions and transform observations into information

**Earth Science Data and Information System (ESDIS)** – Process, archive and distribute Earth Science satellite data and data from airborne and field campaigns (via EOSDIS – the Earth Observing System Data and Information System)

**Global Earth Observation System of Systems (GEOSS)** – Collaborative, international effort to share and integrate Earth observation data. Joint effort with NOAA and the EPA.

**Space Act Agreement** – Collaborative effort with Cray, Inc. to explore development and application of low-latency big data systems

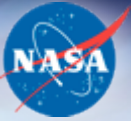
**Planetary Data System (PDS)** – Products from NASA planetary missions

**Multi-Mission Archive at the Space Telescope Science Institute** – Astronomical data archive and tools

**Earth System Grid Federation** – Public archive to support the IPCC Fifth Assessment Report. Collaboration with DOE.



# What does Big Data have to do with you?



If you work in Earth Science,  
you are in the world of Big Data

And it will only get bigger and more complex

NASA is trying to help!!



# **“Big Data”**

What is it?

Why do you care?

## **➤ NASA’s Big Data**

What's out there?

Where is it?

How do you access/acquire it?

# Major Categories of NASA Data



Satellite Remote Sensing Data

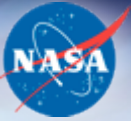
Assimilated Datasets (Validation Data)

Model Output

Climate Projections



# Working with Satellite Remote Sensing Data from NASA



Finding it

Understanding it

Working with it

And most importantly...  
Getting help!!!

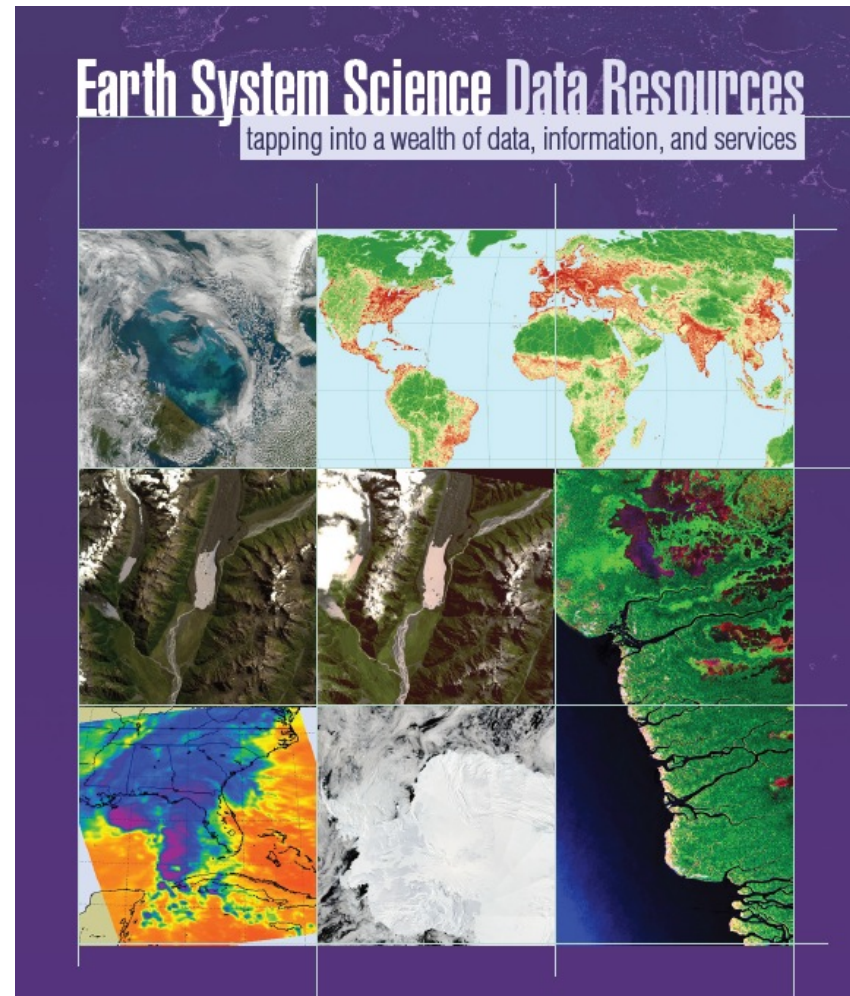
# Start Here!! Your guide to EOSDIS resources



Contains key information about...

- Sensors on NASA spacecraft
- Data and file formats
- NASA's 12 data centers and their data holdings
- Tools for data discovery and access
- Other sources for more detailed information

You will each receive a copy  
... and it is on the web!



<http://earthdata.nasa.gov/library/earth-system-science-data-resources>

# NASA's Earth Science Data Centers



- Alaska Satellite Facility Synthetic Aperture Radar Data Center (ASF SDC)
- Crustal Dynamics Data Information System
- GSFC Earth Sciences Data and Information Services Center (GES DISC)
- Global Hydrology and Resource Center DAAC (GHRC DAAC)
- Langley Research Center Atmospheric Science Data Center (ASDC)
- Land Processes DAAC (LP DAAC)
- MODAPS Level 1 Atmosphere Archive and Distribution System (MODAPS LAADS)
- National Snow and Ice Data Center DAAC (NSIDC DAAC)
- Oak Ridge National Laboratory DAAC (ORNL DAAC)
- Ocean Biology Processing Group
- Physical Oceanography DAAC (PO DAAC)
- Socioeconomic Data and Applications Center (SEDAC)

# The Tools available to you



## The very top level...

- REVERB – A web-based client for discovering and ordering cross-discipline data from all twelve data centers' metadata holdings
- Global Change Master Directory (GCMD) – Allows users to discover and access more than 25,000 Earth science data sets and services
- Search and order within each data center

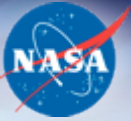


# The Tools available to you

## More detailed tools and services

- ~ 75 tools distributed across the centers for specialized access and manipulation of data sets
- Five categories
  - Search and Order
  - Data Handling
  - Subsetting & Filtering
  - Geolocation, Reprojection & Mapping
  - Data Visualization & Analysis

# Finding data collections



## Key Factors

What physical measurement do you want?

What spatial coverage and resolution do you need?

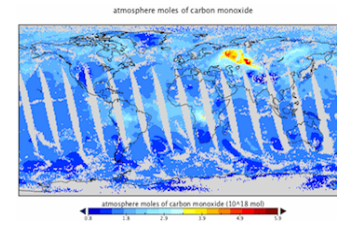
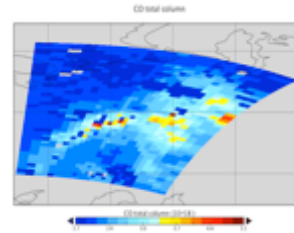
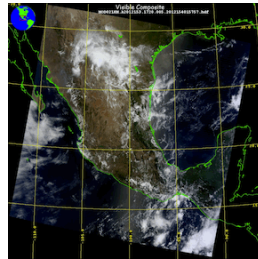
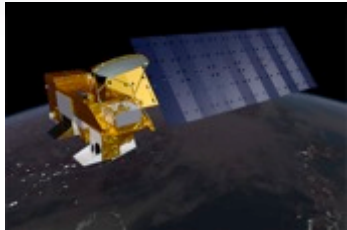
What temporal coverage and resolution do you need?

These will get you started, but you will often find that you need to understand more about how the data were collected and created

Instrument characteristics

Processing levels

# Satellite data processing levels



**Level 0 (L0)**  
raw data  
packets

**Level 1b (L1b)**  
calibrated  
geolocated swaths

**Level 2 (L2)**  
geophysical measurement  
swaths

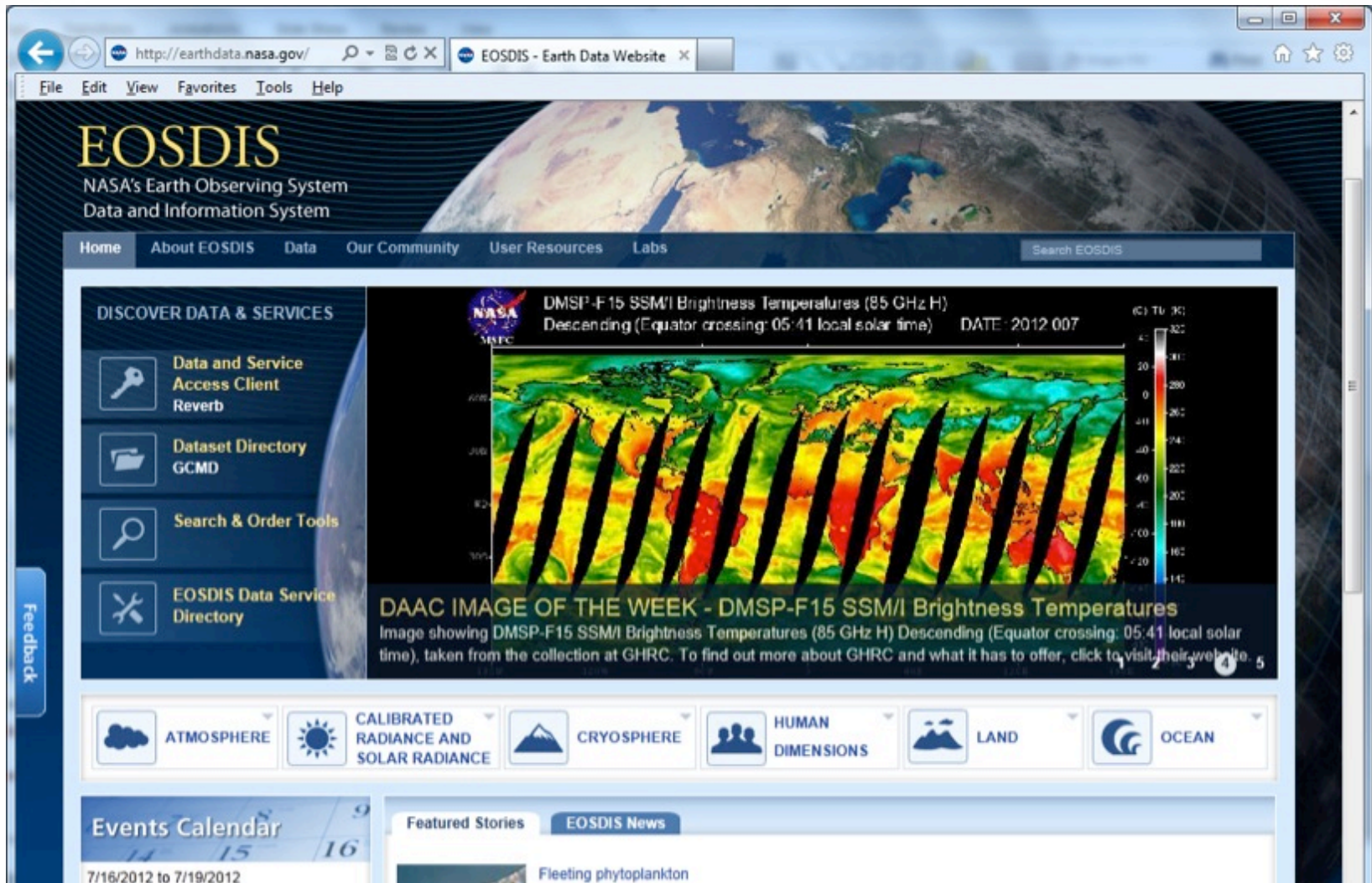
**Level 3 (L3)**  
geophysical measurement  
regular grids

**Level 4 (L4)**  
model output  
regular grids

**EOSDIS Data Center**

# The EOSDIS Portal

<http://earthdata.nasa.gov>



The screenshot shows the EOSDIS Earth Data Website interface. The browser address bar displays <http://earthdata.nasa.gov/>. The website header includes the EOSDIS logo and the text "NASA's Earth Observing System Data and Information System". A navigation bar contains links for Home, About EOSDIS, Data, Our Community, User Resources, and Labs, along with a search bar labeled "Search EOSDIS".

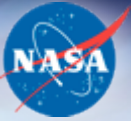
On the left side, a sidebar titled "DISCOVER DATA & SERVICES" lists several options: "Data and Service Access Client Reverb", "Dataset Directory GCMD", "Search & Order Tools", and "EOSDIS Data Service Directory". A "Feedback" button is also visible.

The main content area features a large image of a satellite map showing "DMSP-F15 SSM/I Brightness Temperatures (85 GHz H) Descending (Equator crossing: 05:41 local solar time) DATE: 2012 007". A color scale on the right indicates temperature values from -14 to 32. Below the image, a section titled "DAAC IMAGE OF THE WEEK - DMSP-F15 SSM/I Brightness Temperatures" provides a description of the image and a link to visit the GHRC website.

At the bottom, a horizontal menu categorizes data by domain: ATMOSPHERE, CALIBRATED RADIANCE AND SOLAR RADIANCE, CRYOSPHERE, HUMAN DIMENSIONS, LAND, and OCEAN. Below this, there is an "Events Calendar" for the period 7/16/2012 to 7/19/2012 and a "Featured Stories" section with the title "EOSDIS News" and a sub-story titled "Fleeting phytoplankton".



# Reverb: Access All EOSDIS Data



Reverb | ECHO

http://reverb.echo.nasa.gov/reverb/#utf8=v&spatial\_map=satellite&spatial\_type=rectangle

Untitled presentation - Google Docs    Reverb | ECHO

NASA National Aeronautics and Space Administration

**EOSDIS** NASA's Earth Observing System Data and Information System

**Reverb | ECHO** The Next Generation Earth Science Discovery Tool

EOSDIS Home | Reverb Home | About | Tutorial | Show Help | Shopping Cart (0) | Order Status | Service Request Status | Sign In

### Search Options

- Spatial
- Search Terms
- Temporal
- Platforms & Instruments
- Campaigns
- Processing Levels
- Science Keywords

Save Query  
Clear Criteria

Feedback? Tell us what you think.

Availability

**ASTER GDEM V2 Tutorial**  
2011-10-17 4:00AM (GMT-4:00) to (End Date Not Provided)  
More

Notices

**AMSR-E Instrument Failure**  
2011-10-04 8:58AM (GMT-4:00) to (End Date Not Provided)  
More

Release Information

**Upcoming Features**  
2011-03-24 4:22PM (GMT-4:00)  
An overview of features available in future versions of Reverb.  
More

### Step 1: Select Search Criteria

#### Spatial Search

Bounding Box: e.g. -50.736, 163.477, -11.144, 105.680 (S,E,N,W)    Reset    Clear

Satellite

Click and drag to set a bounding rectangle

Search by ESRI shape file

#### Search Terms

e.g. MODIS Fire AST\_L1A    Clear

#### Temporal Search

START  
YYYY-MM-DD HH:MM:SS  
Clear

END  
YYYY-MM-DD HH:MM:SS  
Clear

\* all times must be specified in GMT

Date Range    Annual Repeating Dates

### Step 2: Select Datasets

Found 2784 datasets. Total Query Time: 0.001s

<input type="checkbox"/>	15 MINUTE STREAM FLOW DATA: USGS (FIFE) Archive Center: ORNL DAAC    Short Name: 15 MINUTE STREAM FLOW DATA: USGS (FIFE)    Version: 0	+/-	0
<input checked="" type="checkbox"/>	2000 Pilot Environmental Sustainability Index (ESI) Archive Center: SEDAC    Short Name: CIESIN_SEDAC_ESI_2000    Version: 1.0	+/-	0
<input checked="" type="checkbox"/>	2001 Environmental Sustainability Index (ESI) Archive Center: SEDAC    Short Name: CIESIN_SEDAC_ESI_2001    Version: 2.0	+/-	0
<input checked="" type="checkbox"/>	2002 Environmental Sustainability Index (ESI) Archive Center: SEDAC    Short Name: CIESIN_SEDAC_ESI_2002    Version: 3.0	+/-	0
<input checked="" type="checkbox"/>	2005 Environmental Sustainability Index (ESI) Archive Center: SEDAC    Short Name: CIESIN_SEDAC_ESI_2005    Version: 4.0	+/-	0

# Global Change Master Directory



The screenshot shows the Global Change Master Directory website. At the top is a header with the title "Global Change Master Directory" and the subtitle "Discover Earth science data and services". Below this is a navigation bar with links: Home, Data Sets, Data Services, Portals, Add to GCMD, What's New, Participate, CEOS IDN, and About GCMD. The main content area is divided into two columns. The left column, titled "Find Data", lists various Earth science categories with corresponding icons: Agriculture (forest science, soils ...), Atmosphere (precipitation, air quality ...), Biosphere (ecosystems, vegetation...), Biological Classification (animals/invertebrates, plants...), Climate Indicators (air temperature, drought ...), Cryosphere (frozen ground, sea ice ...), Human Dimensions (land use, population ...), and a section for Data Centers, Projects, Instruments / Sensors, and Platforms / Sources. The right column, titled "Find Data Services", lists various data services: Data Analysis and Visualization, Data Management / Data Handling, Education / Outreach, Environmental Advisories, Hazards Management, Metadata Handling, Models, Reference and Information Services, and Web Services. At the bottom of the right column is a "Data Services Text Search" box. The footer of the website includes the CEOS International Directory Network logo and a search bar.

**Global Change Master Directory**  
Discover Earth science data and services

Links | FAQ | Contact Us | Site Map

Home | Data Sets | Data Services | Portals | Add to GCMD | What's New | Participate | CEOS IDN | About GCMD

*Find Data*

- Agriculture**  
forest science, soils ...
- Atmosphere**  
precipitation, air quality ...
- Biosphere**  
ecosystems, vegetation...
- Biological Classification**  
animals/invertebrates, plants...
- Climate Indicators**  
air temperature, drought ...
- Cryosphere**  
frozen ground, sea ice ...
- Human Dimensions**  
land use, population ...
- Data Centers** **Projects**
- Instruments / Sensors**
- Platforms / Sources**

**Land Surface**  
erosion, topography ...

**Oceans**  
ocean temperature , salinity ...

**Paleoclimate**  
ice cores, land records ...

**Solid Earth**  
geochemistry, seismology ...

**Spectral / Engineering**  
radar, visible imagery ...

**Sun-Earth Interactions**  
auroras, solar activity ...

**Terrestrial Hydrosphere**  
ground water, water quality ...

**Map / Date Search**  
or  
**Location Name Search**

*Find Data Services*

- [Data Analysis and Visualization](#)
- [Data Management / Data Handling](#)
- [Education / Outreach](#)
- [Environmental Advisories](#)
- [Hazards Management](#)
- [Metadata Handling](#)
- [Models](#)
- [Reference and Information Services](#)
- [Web Services](#)

[Data Services Text Search](#)

[Find](#) [Data Only](#)

**CEOS**  
International Directory Network

*Data Set Text Search*

# Drilling down into Land Surface data

Refine by Category

Refine by Full text

[Show All Titles for LAND SURFACE](#) (4672)

## [EROSION/SEDIMENTATION](#) (459)

degradation , entrainment , erosion , landslides , sediment chemistry ...

## [FROZEN GROUND](#) (309)

active layer , cryosols , ground ice , periglacial processes , permafrost ...

## [GEOMORPHOLOGY](#) (298)

coastal landforms/processes , eolian landforms/processes , fluvial landforms/processes , glacial landforms/processes , karst landforms/processes ...

## [LAND TEMPERATURE](#) (425)

land heat capacity , land surface temperature , skin temperature ...

## [LAND USE/LAND COVER](#) (1305)

land cover , land productivity , land resources , land use classes ...

## [LANDSCAPE](#) (345)

landscape ecology , landscape management , landscape patterns , landscape processes , reclamation/revegetation/restoration ...

## [SOILS](#) (1645)

calcium , carbon , cation exchange capacity , denitrification rate , electrical conductivity ...

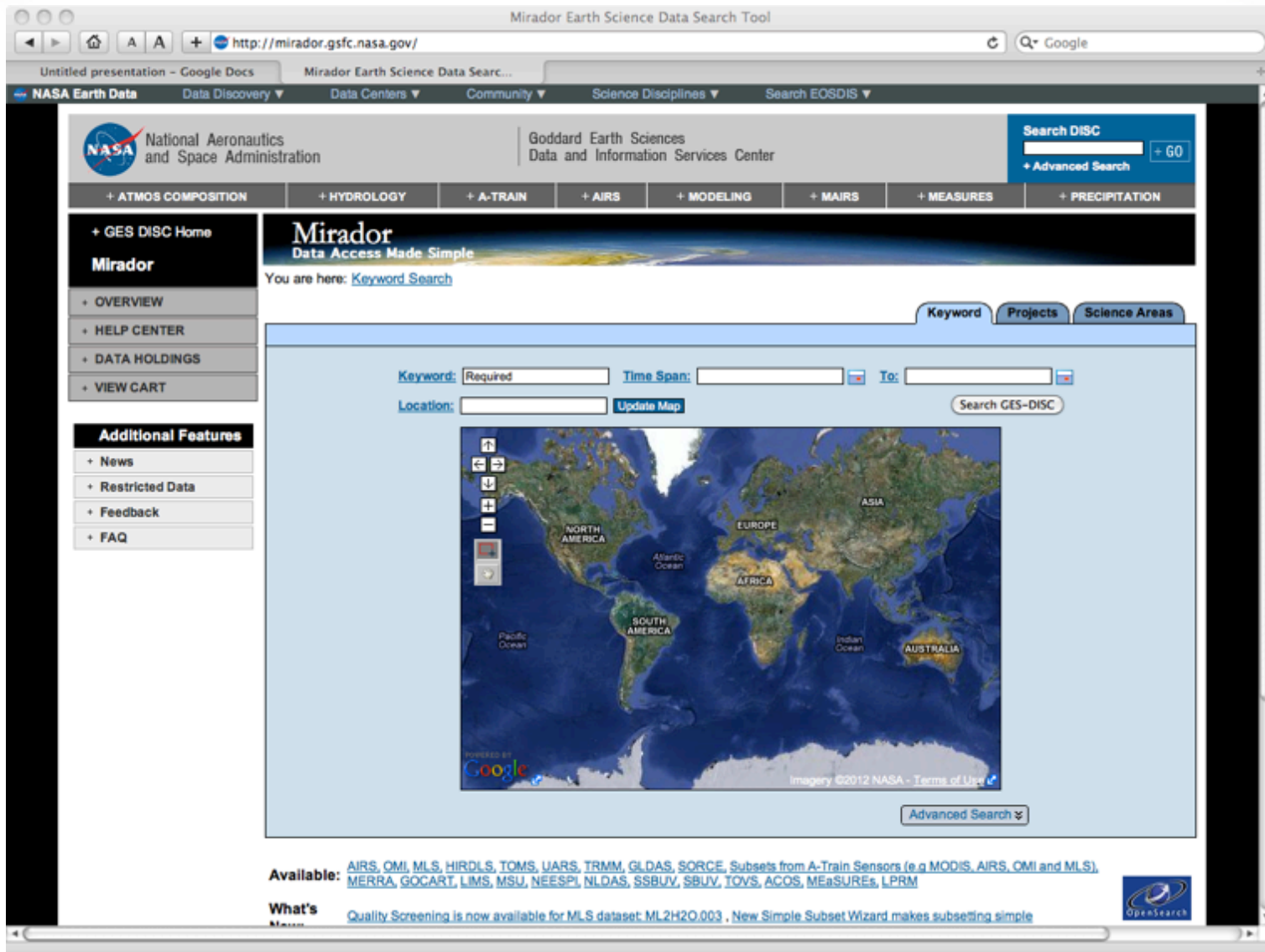
## [SURFACE RADIATIVE PROPERTIES](#) (479)

albedo , anisotropy , emissivity , reflectance , thermal properties ...

## [TOPOGRAPHY](#) (1534)

contours , landforms , surface roughness , terrain elevation , topographic effects ...

# Mirador: Search Goddard Earth Sciences Data and Information Services Center (GES DISC)



The screenshot displays the Mirador Earth Science Data Search Tool interface within a web browser. The browser's address bar shows the URL <http://mirador.gsfc.nasa.gov/>. The page features a NASA logo and the text "National Aeronautics and Space Administration" and "Goddard Earth Sciences Data and Information Services Center". A search bar labeled "Search DISC" is located in the top right corner, with a "GO" button and a link to "Advanced Search". Below the search bar, a horizontal menu lists various data categories: + ATMOS COMPOSITION, + HYDROLOGY, + A-TRAIN, + AIRS, + MODELING, + MAIRS, + MEASURES, and + PRECIPITATION. On the left side, a sidebar contains links to "GES DISC Home", "Mirador", "OVERVIEW", "HELP CENTER", "DATA HOLDINGS", "VIEW CART", and "Additional Features" (News, Restricted Data, Feedback, FAQ). The main content area is titled "Mirador Data Access Made Simple" and shows the user's current location as "Keyword Search". It includes a search form with fields for "Keyword" (with a "Required" label), "Time Span", "To", and "Location", along with an "Update Map" button and a "Search GES-DISC" button. Below the search form is a world map showing continents and oceans, with labels for NORTH AMERICA, SOUTH AMERICA, EUROPE, ASIA, AFRICA, and AUSTRALIA. The map is powered by Google and includes a copyright notice for Imagery ©2012 NASA - Terms of Use. At the bottom, there is a section for "Available" data sources, listing various sensors and datasets, and a "What's New" section mentioning quality screening and a new subset wizard.

Available: [AIRS](#), [OMI](#), [MLS](#), [HIRDLS](#), [TOMS](#), [UARS](#), [TRMM](#), [GLDAS](#), [SORCE](#), Subsets from A-Train Sensors (e.g. [MODIS](#), [AIRS](#), [OMI](#) and [MLS](#)), [MERRA](#), [GOCART](#), [LIMS](#), [MSU](#), [NEESPI](#), [NLDAS](#), [SSBUV](#), [SBUV](#), [TOVS](#), [ACOS](#), [MEASURES](#), [LPRM](#)

What's New: Quality Screening is now available for MLS dataset: [ML2H2O.003](#) , New Simple Subset Wizard makes subsetting simple



# Simple Subset Wizard



Simple Subset Wizard (SSW)

http://disc.sci.gsfc.nasa.gov/SSW/

Google

Untitled presentation - Google Docs

Simple Subset Wizard (SSW)

NASA National Aeronautics and Space Administration

EOSDIS NASA's Earth Observing System Data and Information System

Home About EOSDIS Data Our Community User Resources Labs Search EOSDIS

EOSDIS Home

**SIMPLE SUBSET WIZARD (SSW)** [V1.05 RELEASE NOTES](#)

Enter values for the Date Range and (optionally) the Spatial Bounding Box to search for data sets; those criteria will also be used when data sets are subsetted by Date Range and Spatial Region.

Enter keywords or click the "Select Data Sets" button.

Data Set Keyword(s)

Enter dates as YYYY-MM-DD or use the calendars.

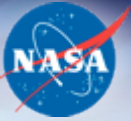
Date Range   to

Enter South, West, North, East coordinates or use the map.

Spatial Bounding Box

[Report a Problem with the Simple Subset Wizard](#)

# Using Data



## Data Formats

- HDF4, HDF5: advanced data formats requiring an API or tool
- HDF-EOS, HDF-EOS5: HDF with standard structures for geolocation
- netCDF: advanced data formats requiring an API or tool
- netCDF/CF-1: netCDF with standards for geolocation, time and vertical coordinates
- Some flat binary formats (ASCII and GRIB) and codes like BUFR

# netCDF Tools with GUI

## Panoply

- <http://www.giss.nasa.gov/tools/panoply>
- quick-view, shallow learning curve
- cross-platform
- reads some HDF (L2), lots of OPeNDAP

## IDV (integrated data viewer)

- <http://www.unidata.ucar.edu/software/idv/>
- sibling product is McIDAS-V
- More flexible than Panoply
- A little harder to learn
- Reads many other formats, including OPeNDAP, but NOT HDF

# non-GUI netCDF Tools



## nco (netcdf command operators)

- Command-line software for doing math on netCDF files
- Useful utilities for manipulating (editing) netCDF files

## ncl (netcdf command language)

- More geared toward plotting

Ferret: <http://ferret.wrc.noaa.gov/Ferret/>

- Geared toward plotting

## Gridded Analysis and Display System (GrADS)

- <http://www.iges.org/grads/>
- Only works well with L3-L4 data
- Can read OPeNDAP
- Can read HDF with more effort



# HDF Tools



## HDFView

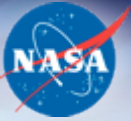
- <http://www.hdfgroup.org/hdf-java-html/hdfview/>

hdp (HDF4), h5dump (HDF5)

- Comes with HDF libraries
- Extracts data from HDF files into ASCII or flat binary

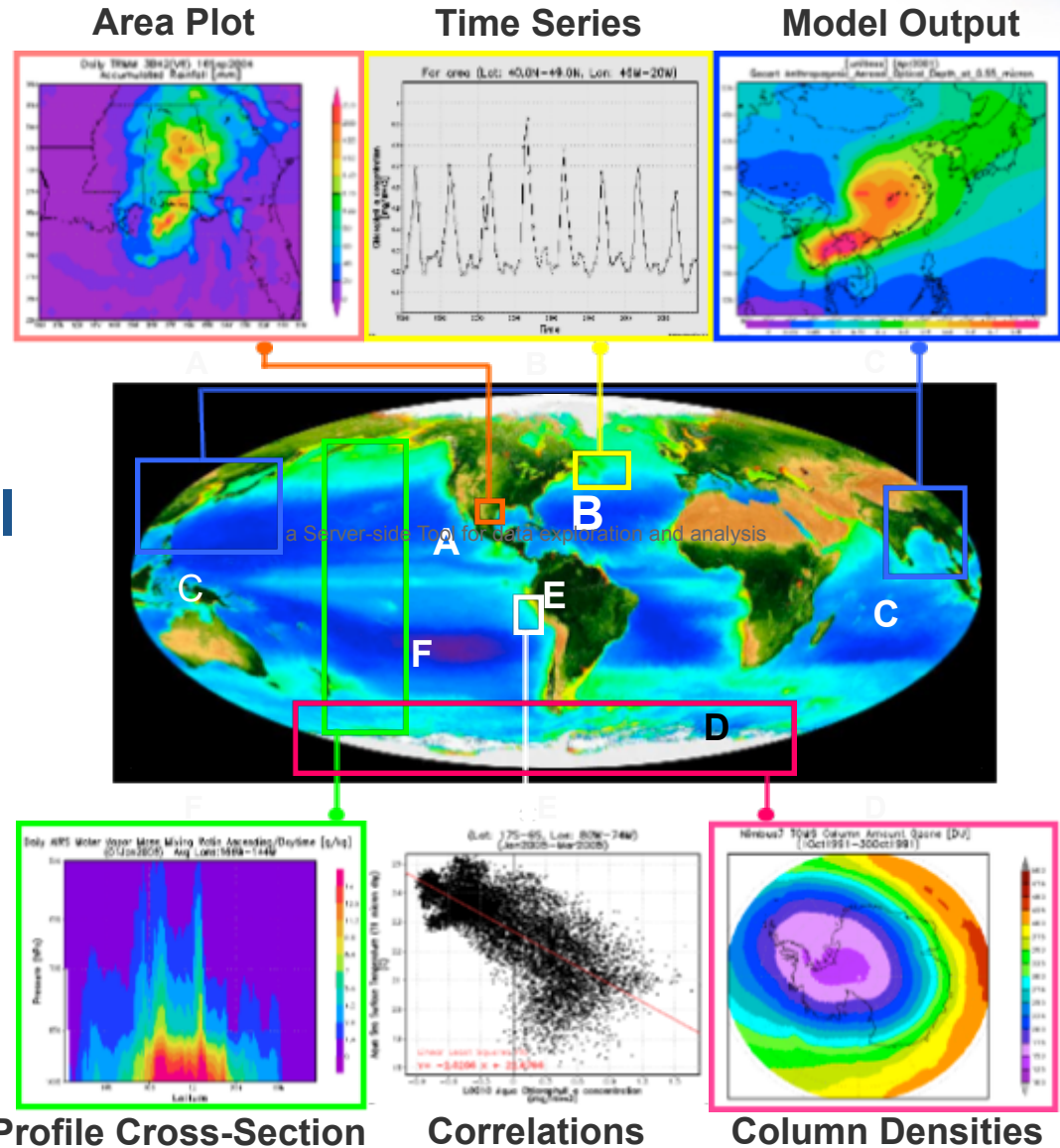
HDF data served through OPeNDAP “looks” like netCDF data to most clients

# Visualizing and Analyzing Data



**Giovanni**

**A Server-side Tool  
for data  
exploration and  
analysis**



# Giovanni Interface

Select Area of interest

Select Variables

Select Time Period

Select Plot type

Generate Visualization

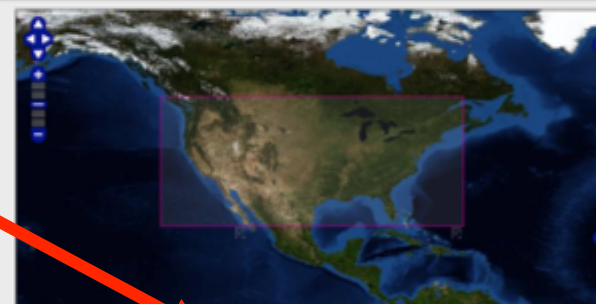
**Giovanni** - The Bridge Between Science and Data

**Giovanni Air Quality**  
EPA AIRNOW PM2.5, MODIS and OMI Measurements for Air Quality Applications

This Giovanni interface provides gridded 1x1 degree EPA AIRNOW surface PM2.5 concentrations over the continental U.S. to allow collocation with satellite aerosol products for analysis of U.S. air quality. The PM2.5 data (acquired through DataFeed) are the average of hourly concentrations between 152 and 222 to maximize temporal collocation with the MODIS Terra and Aqua overpass times over the U.S. All satellite products are at 1x1 degree resolution.  
**NOTE: All satellite aerosol products in this instance are column measurements. Satellite detected aerosols may or may not be located at the surface. PM2.5 concentrations are in situ surface measurements.**

Select Constraints:

Spatial



Press 'Shift' key to draw a zoom rectangle. Mouse wheel zooms in and out. Enter selection coordinates below.

☐ Pan ☐ Draw Box West: -129 North: 51 South: 24 East: -66 Update

Parameters

Display: ☒ Data Product Info ☐ Units

☐ AIRNOW\_PM2.5 (EPA monitor) - 2000/10/01 - 2000/10/01

☐ Fine Particulate Matter - PM2.5

☐ MODIS\_Terra (MODIS) - 2000/10/01 - 2000/10/01

☐ Aerosol Optical Depth at 550 nm

☐ Aerosol Small Mode Optical Depth

☐ Cloud Optical Depth - Total (QA-w)

☐ Cloud Top Pressure (Day only)

☐ Cloud Top Temperature (Day only)

☐ MODIS\_Terra (MODIS) - 2000/10/01 - 2000/10/01

☐ Aerosol Optical Depth at 550 nm

☐ Aerosol Small Mode Optical Depth

☐ Cloud Optical Depth - Total (QA-w)

☐ Cloud Top Pressure (Day only)

☐ Cloud Top Temperature (Day only)

☐ CASSE/OMI (OMI) - 2000/10/01 - 2000/10/01

☐ Aerosol Absorption Optical Depth at 550 nm

☐ Aerosol Extinction Optical Depth at 550 nm

☐ OMI\_Terra (OMI) - 2000/10/01 - 2000/10/01

☐ Aerosol Index

Begin Date Year 2000 Month Sep Day 1

End Date Year 2000 Month Sep Day 1

Select Visualization:

Lat-Lon map, Time-averaged

Generate Visualization Reset

**Animation**

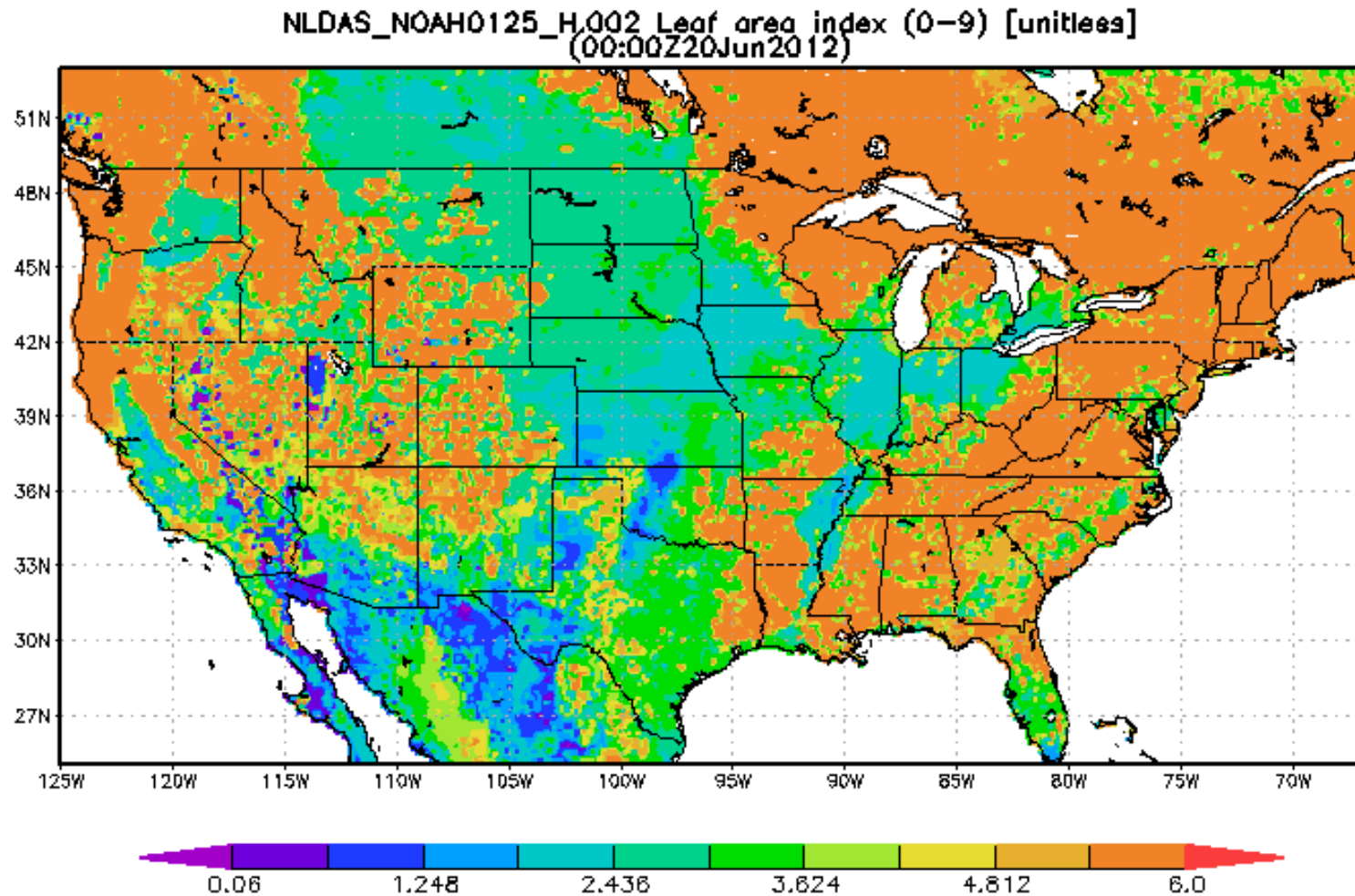
- Overlay of Lat-Lon Maps
- Latitude-Time Hovmöller Diagram
- Longitude-Time Hovmöller Diagram
- Lat-Lon map, Time-averaged
- Correlation map**
- Lat-Lon map of time-averaged differences
- Scatter plot
- Scatter plot, Time-averaged
- Time series
- Time series, Area-averaged differences
- Time series, Area statistics
- Correlation map

NASA High End Computing Capability

2012 Summer Short Course for Earth Systems

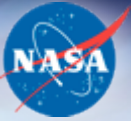
Responsible NASA Official: Steven J. Kempner@nasa.gov  
Web Curator: Stephen W. Bernick web-contact@nasa.gov  
Privacy Policy and Important Notices

# E.g., NLDAS Leaf Area Index





# Giovanni Details



<http://giovanni.gsfc.nasa.gov>

- Mostly Level 3 data
- Mostly data from GES DISC

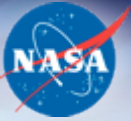
## Several available functions

- Time averaging, spatial averaging, regression

## Land Data in Giovanni

- Monsoon Asia Integrated Regional Study (MAIRS)
- Northern Eurasia Earth Science Partnership Initiative (NEESPI)
- North American Land Data Assimilation System (NLDAS)
- Global Land Data Assimilation System (GLDAS)

# Model and Assimilation Data



- MERRA – Using 30 years of satellite data and the most advanced modeling and data assimilation techniques to create a reference data set
- ESGF – Consolidating access to the massive amount of model data being created for the next IPCC Assessment Report (AR5)



# MERRA

## Modern Era Retrospective Analysis for Research and Applications

What is MERRA?

NASA's ambitious effort to extract new information from 30 years of satellite observations using GEOS-5 Data Assimilation System

- New analysis technique for 4D representation of the state of the atmosphere
- Exploits NASA and NOAA rich satellite data holdings, additional satellite retrieval products, and conventional observations
- Provides data at higher spatial and temporal resolution

# MERRA

## Modern Era Retrospective Analysis for Research and Applications

### Why is MERRA Important?

Provides invaluable data source for model assessment, climate & weather research, and hydrologic studies

- **Model Validation** – re-analysis datasets provide an improved representation of observational data to support assessment of model forecast skill
- **Scientific Research** – re-analysis datasets provide an improved derivation of geophysical parameters from observational data to support scientific study



# MERRA

## Modern Era Retrospective Analysis for Research and Applications

How do you get MERRA data?

NASA/GSFC Global Modeling and Assimilation Office (GMAO)

- Understanding comparative merits of MERRA  
<http://gmao.gsfc.nasa.gov/research/merra/prequel/view.php>
- Finding and manipulating MERRA data with OpenDAP  
[http://gmao.gsfc.nasa.gov/pubs/docs/da\\_Silva380.pdf](http://gmao.gsfc.nasa.gov/pubs/docs/da_Silva380.pdf)

NASA Goddard Earth Sciences Data & Information Services  
Center (GES DISC)

- Accessing MERRA data via GES DISC  
<http://daac.gsfc.nasa.gov/>
- Re-analysis and observing system data manipulation with Giovanni  
<http://giovanni.gsfc.nasa.gov/>

# Accessing MERRA via GES DISC



The screenshot displays the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) website. The header includes the NASA logo, the center's name, and a search bar with a 'GO' button. Below the header is a navigation bar with tabs for different data categories: ATMOS COMPOSITION, HYDROLOGY, A-TRAIN, AIRS, MODELING, MAIRS, and PRECIPITATION. The main content area is divided into several sections. On the left, there's a 'GES DISC' sidebar with links to 'ABOUT US', 'DATA HOLDINGS', and 'SERVICES'. Below this is an 'Additional Features' section with links to 'News', 'Image Gallery', 'Science Focus', 'Technology Lab', 'Publications', 'Citing Our Data', 'FAQ', and 'Science Team Portal (Restricted Access)'. The main content area features a 'WHAT'S HOT' section with three highlighted items: 'MERRA Meteorological Data Now Available from the GES DISC', 'Giovanni Air Quality Instance now available', and 'Hydrology DISC supports GLDAS data product'. Each item includes a small thumbnail image and a 'Read More' link. To the right of the main content is a 'EARTH MEASUREMENTS' sidebar with links to 'ATMOS COMPOSITION', 'A-TRAIN DATA DEPOT', 'HURRICANES', 'HYDROLOGY', 'OCEANS', and 'PRECIPITATION'. Below this is a 'CURRENT MISSION DATA' sidebar with links to 'AIRS', 'AURA', 'SORCE', and 'TRMM'. At the bottom right, there are three more featured sections: 'Giovanni Online Visualization and Analysis', 'Mirador Data Access Made Simple', and 'NASA Imagery in Virtual Globes'. The footer of the page contains a 'LATEST NEWS' section with a headline dated 11.09.09 about the utilization of Giovanni for research.

NASA National Aeronautics and Space Administration | Goddard Earth Sciences Data and Information Services Center

Search DISC  + GO  
+ Advanced Search

+ ATMOS COMPOSITION + HYDROLOGY + A-TRAIN + AIRS + MODELING + MAIRS + PRECIPITATION

**GES DISC**  
Goddard Earth Sciences Data and Information Services Center  
Your source for earth science data and information

+ ABOUT US  
+ DATA HOLDINGS  
+ SERVICES

**Additional Features**  
+ News  
+ Image Gallery  
+ Science Focus  
+ Technology Lab  
+ Publications  
+ Citing Our Data  
+ FAQ  
+ Science Team Portal (Restricted Access)

**REMOTE SENSING DATA**  
Make Selection ▼

**OTHER PORTALS**  
Make Selection ▼

**GES DISC**  
Goddard Earth Sciences Data and Information Services Center  
Your source for earth science data and information

**WHAT'S HOT**

**MERRA Meteorological Data Now Available from the GES DISC**  
The Modern Era Retrospective-analysis for Research and Applications (MERRA) data production is underway. Multi-year segments of the meteorological analyses from each of the 3 production streams are now available online at the GES DISC.  
  
[+ Read More](#)

**Giovanni Air Quality Instance now available**  
A new Giovanni instance dedicated to air quality related data is now available. The instance includes global aerosol and cloud data from MODIS, global aerosol data from OMI, and the AIRNow Fine Particulate Matter (PM2.5) ground-based monitoring product for the continental United States.  
  
[+ Read More](#)

**Hydrology DISC supports GLDAS data product**  
The new Hydrology DISC (HDISC) portal provides access and support for the Global Land Data Assimilation Systems (GLDAS).  
  
[+ Read More](#)

**EARTH MEASUREMENTS**  
+ ATMOS COMPOSITION  
+ A-TRAIN DATA DEPOT  
+ HURRICANES  
+ HYDROLOGY  
+ OCEANS  
+ PRECIPITATION

**CURRENT MISSION DATA**  
+ AIRS  
+ AURA  
+ SORCE  
+ TRMM

**Giovanni**  
Online Visualization and Analysis

**Mirador**  
Data Access Made Simple

**NASA Imagery in Virtual Globes**

**LATEST NEWS**  
**11.09.09 - Utilization of Giovanni for research expands in global science community**  
Scientific use of the NASA GES DISC Interactive Online Visualization And Analysis Infrastructure, Giovanni, is proliferating. The number of peer-reviewed research papers utilizing Giovanni surpassed 100 in 2009; in 2009 alone, 46 papers in which Giovanni was used have already been published.

# NASA Model and Observation Data on the Earth System Grid (ESG)



Earth System Grid Federation – a partnership of climate modeling centers

- Provide secure web-based distributed access to Coupled Model Intercomparison Project Phase 5 (CMIP5) model data that will be the basis for IPCC AR5
- Ability to search, download analyze and visualize this distributed data through one common web interface
- NCCS ESG node provides access to
  - Model data produced by the GMAO and GISS
  - CMIP5 data from NOAA and Center for Ocean-Land-Atmosphere Studies (COLA) produced with the National Centers for Environmental Prediction (NCEP) model
  - Selected MERRA datasets

# NCCS ESG Node Home Page



[Home](#) [Search](#) [Tools](#) [Login](#)



Welcome to the ESGF Index for the NCCS at NASA/GSFC

Welcome to the new CMIP5 distributed archive. Our new ESGF peer-to-peer (P2P) enterprise system is now the official site for CMIP5 model output. The old [gateways](#) will remain active and output from all models will continue to be available until the end of July 2012. Please send e-mail to [esgf-user@lists.llnl.gov](mailto:esgf-user@lists.llnl.gov) to report bugs and provide feedback.



## Quick Search

Keyword:

Advanced Search (Category, Geospatial, Temporal, and more)...



## Peer Nodes

- [ANL Node](#)
- [CMCC Node](#)
- [DKRZ Node](#)
- [DKRZ CMIP5 Node](#)
- [IPSL Node](#)
- [NASA-GSFC Node](#)
- [NASA-JPL Node](#)
- [ORNL Node](#)



## About esgf.nccs.nasa.gov

The NASA Center for Climate Simulation (NCCS) offers an integrated set of supercomputing, visualization, and data services to enhance weather and climate prediction capabilities. It serves hundreds of users across the U.S., including the [Goddard Institute for Space Studies \(GISS\)](#) and the [Global Modeling and Assimilation Office \(GMAO\)](#). The NCCS ESGF Node serves [CMIP5](#) model simulation results from GISS, GMAO, and climate research colleagues at NOAA's [NCEP](#) and [COLA](#). It also provides access to specific [MERRA reanalysis data](#) and observational data ([CERES-EBAF](#), and [TRMM](#)). These observational datasets are provided as part of an experimental activity to increase the usability of NASA satellite observational data for the model and model analysis communities. These are not



## Resources



### Quick Links

- [Create Account](#)
- [MyProxyLogin](#)
- [Expert Search \(XML\)](#)
- [Web Script Generator](#)
- [ESGF aggregated RSS feed](#)
- [Contact ESGF](#)



### Instructions

- [ESGF Full User Guide](#)
- [Search Help](#)
- [Search Controlled Vocabulary](#)
- [Web Scripts FAQ](#)



# And finally - NEX Data Holdings

These are datasets easily available to researchers with accounts on the NEX systems

Dataset	Variable	Volume	Location
NARR	Multiple	2.6TB	Pleiades, Lou
ASTER	DEM	256GB (gz)	Pleiades, Sandbox
CARBONTRACEKR	Multiple	701GB (gz)	Pleiades, Lou
CMIP	Multiple	2TB	Pleiades, Lou
CRU-NCEP	Multiple	1.3TB	Pleiades, Sandbox, Lou
GIMMS	Multiple	13GB	Pleiades, Lou
GTOPO30	Multiple	6GB	Pleiades, Sandbox, Lou
LANDSAT	Multiple	240TB (gz)	Pleiades, subset on Sandbox, Lou
MODIS	Multiple	30TB	Pleiades, subset on Sandbox, Lou
MSTMIP	Multiple	4TB	Pleiades, Sandbox, Lou
NACP	Multiple	500BG	Pleiades, Lou
NED	Multiple	26GB	Pleiades, Sandbox, Lou
PRISM	Multiple	1TB	Pleiades, Lou
SRTM	DEM	100GB	Pleiades, Sandbox, Lou
TOPS	Multiple	20TB	Lou
TOPS-CLIM	Multiple	10TB	Sandbox, Lou
MERRA	Multiple	59GB	Sandbox, Lou
NCDC	Multiple	17GB	Pleiades, Sandbox, Lou
NCEP	Multiple	21GB	Sandbox, Lou
TRMM	Multiple	1GB	Sandbox, Lou
OMI	Multiple	1GB	Sandbox, Lou
STATSGO	Multiple	1GB	Sandbox, Lou
WELD	Surface Reflectances	20TB	Pleiades
NADF3	Multiple	100GB	Pleiades, Lou
SIMS	Multiple	7TB	Sandbox, Lou



# Thank You!!!

2012 Summer Short Course for Earth System  
Modeling and Supercomputing

# Getting data files

Search by time, space, and data collection attributes  
(satellite, instrument, processing level)

All of EOSDIS: Reverb, <http://reverb.echo.nasa.gov>

One stop shopping, but

Slower, more cumbersome for large data searches

Individual data centers

<http://earthdata.nasa.gov/data/data-tools/search-and-order-tools>

Optimized for the data center's data

Simple Subset Wizard: <http://disc.gsfc.nasa.gov/>

SSW

Covers 10 of 12 datasets

Highly simplified user interface

*Hint:* try SSW first, then other search tools if you don't find what you need

# Getting data files (con't)



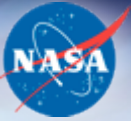
Search tools will provide access to online URLs or ability to order data

Most data centers offer *dataset-specific* additional services

- Subsetting (spatial, by variable)
- MODIS Reprojection Tool
- Data Quality Screening Service (AIRS, MODIS L2)
- Reformat to netCDF
- Simple Subset Wizard offers some of these services
- OPeNDAP, Web Coverage Service, Web Map Service



# Caveat #1: PAY ATTENTION TO DATA QUALITY!!



Most Level 1 and 2 data has quality control parameters

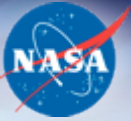
- Also, some L3, e.g., MODIS Land Surface Temperature

Failure to filter data according to quality flags can result in significant bias in your results

E.g., [Data Quality Screening Service](#)

- Only available for certain L2 atmospheric products at this point

## Caveat #2



# Daily Level 3 data is not a daily “average”!

## The Diurnal Sampling Bias

Technically each L3 grid cell value is the average of L2 pixels falling in that cell for that day,  
**BUT...**

In reality, it has contributions from only one or two satellite overpasses during the day\*

- therefore about a 100-minute window during the day
- roughly the SAME 100-minute window every day for sun-synchronous satellites
- so averaging over weeks, months, years does not eliminate the Diurnal Sampling Bias